DOCUMENT RESUME

ED 463 325 TM 033 757

AUTHOR Shindler, John V.
TITLE Examining the Soundness of Two Collaborative Assessment Practices in Teacher Education Courses.
PUB DATE 2002-04-00
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cooperation; Educational Assessment; *Evaluation Methods; Focus Groups; Higher Education; Interviews; *Preservice Teacher Education; *Preservice Teachers; Reliability; *Student Evaluation; Validity
IDENTIFIERS *Collaborative Evaluation

ABSTRACT

Most often, new teachers default to the pedagogical practices they themselves were exposed to as teacher candidates. If teacher education programs are to promote the value of collaboration, they must teach and model collaborative pedagogy within their programs. This study is a qualitative examination of the soundness of two forms of collaborative assessment within teacher education courses: collaborative or group examinations and a system of collaborative interactive roundtable presentations. The construct of soundness is defined within a four-dimensional framework consisting of validity, reliability, efficiency, and effect on the learner. Subjects (n=45, 46, and 248) were students in required methods courses. Data came from participant surveys, focus group interviews, and instructor observation of participants. Results of the study suggest that these collaborative assessment methods compare favorably on all four dimensions of soundness. Conventional wisdom would call into question the ability of these methods to achieve reliable measurements and differentiation of student performance and their efficiency, but participant surveys rated these methods more highly for each of these measures. Participants experienced a greater degree of critical thinking, motivation to prepare, enjoyment of the assessment process, and a better relationship with classmates. Three appendixes contain a definition of assessment soundness, advice on incorporating collaborative examinations into a course, and advice on incorporating a roundtable presentation. (Contains 16 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 325

Examining the Soundness of Two Collaborative Assessment Practices in Teacher Education Courses

John V. Shindler, Ph.D.
Division of Curriculum and Instruction
Charter College of Education
California State University, Los Angeles
jshindl@calstatela.edu

J. Shindler

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM033757

A paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April, 2002

BEST COPY AVAILABLE

Examining the Soundness of Two Collaborative Assessment Practices in Teacher Education Courses

Abstract

Most often new teachers default to the pedagogical practices that they themselves were exposed to as teacher candidates. This point was emphasized in a 1997 Report by NCATE (National Council for Accreditation of Teacher Education), in which they stated, "Today's teacher candidates will teach tomorrow as they are taught today (p.1)." This methodological reproduction suggests an elevated need for those of us in teacher education to model both sound as well as innovative practice. While the field of educational assessment has produced much innovation in the past decade, most assessment in teacher education is still primarily *individualistic*. If teacher education programs are to promote the value of collaboration within their candidates, they must teach and model collaborative pedagogy within their programs. The reticence for using more *collaboratively* structured assessment methods may be that they are seen as less sound.

This study is a qualitative examination of the soundness of two forms of collaborative assessment within teacher education courses. The forms of assessment being investigated are 1) collaborative or group exams, and 2) a system of collaborative, interactive roundtable presentations. The construct of *soundness* is defined within a four-dimensional framework consisting of validity, reliability, efficiency, and effect on the learner. Subjects (N=45, 46, 248) were members of required methods courses. Data consisted of participant surveys, focus group interviews, and instructor participant observation. The results of the study suggest that these collaborative assessment methods compared favorably on all 4 dimensions of soundness. While conventional wisdom would call into question these method's ability to achieve reliable measurements and differentiation of student performances as well as the ability to be performed as efficiently as more traditional methods of assessment, participant surveys rated collaborative methods slightly higher on each of these areas. Moreover, the data suggested that the benefits experienced by the participants taking part in the collaborative methods were significant. Participants experienced a greater degree of critical thinking, motivation to prepare, enjoyment of the assessment process, and relationship with classmates, while reporting that they learned more in the collaborative assessment conditions. A discussion of findings and directions for how collaborative assessment might be implemented into a course are included in the paper.

Examining the Soundness of Two Collaborative Assessment Practices in Teacher Education Courses

Most often new teachers default to the pedagogical practices that they themselves were exposed to as teacher candidates. This point was emphasized in a 1997 Report by NCATE (National Council for Accreditation of Teacher Education), in which they stated, "Today's teacher candidates will teach tomorrow as they are taught today (p.1)." This methodological reproduction suggests an elevated need for those of us in teacher education to model both sound as well as innovative practice. While the field of educational assessment has produced much innovation in the past decade, most assessment in teacher education is still primarily *individualistic*. Current standards from the paramount professional societies in teacher education including NCATE, INTASC, and NBPTS hold *collaboration* skills and dispositions as critical to a well-prepared teacher. For example, INTASC Principle #7, Disposition, #3, states, "The teacher values planning as a collegial activity." If teacher education programs are to promote the value of collaboration within their candidates they must teach and model collaborative pedagogy within their programs. The reticence for using more collaboratively structured assessment methods may be that they are seen as less sound.

This study is a qualitative examination of the soundness of two forms of collaborative assessment within graduate teacher education courses at two large state universities with large teacher education programs. The forms of assessment being investigated are 1) collaborative or group exams, and 2) a system of collaborative interactive roundtable presentations. The construct of *soundness* is defined within a four-dimensional framework consisting of validity, reliability, efficiency, and effect on the learner. Collaborative assessment is rarely used in teacher education and even less outside of education (Antony, 1994). The reticence is likely a result of both its unfamiliarity and the fear that it is not as *sound* as more traditional forms. This study examines each of these concerns, and explores the technical requirements of collaborative assessment usage and compares its soundness to more common methods.

In their limited application, collaborative exams have been shown to improve content retention, promote higher level thinking (Stearns, 1996; Yuretich, Khan, & Leckie, 2001), and increase the overall enjoyment of the course (Stearns, 1996). Interactive presentation formats have been shown to have a similar set of effects (Hermann, 1995; MacDonald, 1989; Schumm, 1995). The collaborative element of the assessments seems to promote a more thoughtful level of processing and more creative work (Bohde, 1996). Moreover both methods seem to provide a potentially more authenticity context, inasmuch as "good teachers" have a greater tendency to plan collaboratively (Fullan, 1993).

THEORETICAL FRAMEWORK FOR SOUNDNESS

This study incorporates a four-dimensional theoretical framework for soundness that has been shown to be conceptually as well as practically robust (Shindler, Yang, Nephew & Keen, 2000). Within this framework, any assessment practice can be considered sound to the degree that it possesses validity, reliability, efficiency, and has a positive effect on its users. Validity is defined by the degree to which a method measures the most important concepts, matches the content covered, and is the best-suited form of methodology to capture the desired learning. Reliability could be characterized by the degree to which a method can obtain an accurate representation of the learning, both among raters (or hypothetical rates) and across multiple performances. Efficiency deals with how "doable" an assessment method is, and how well it can be performed without either taking time away from other teaching and/or other learning. The area related to the effect on the learner could also be considered what has been termed "consequential validity," but is dealt with as a separate consideration here. This dimension includes the motivational, psychological and epistemological affects the assessment has on any learner and/or the class as a whole. (See Appendix A for working definition of soundness provided students)

METHODS

The Two Study Assessment Conditions

1. Cooperative Group Exams

Assessment Procedure:
Condition A: In this exam format, students are allowed to work together to develop their response to written exam prompts, but each student's exam is evaluated individually. Students are allowed to choose their own groups, and because there should have been a great deal of cooperative class work to this point, they are familiar with one another and are in a good position to purposefully select a team. Opting to work alone is allowed at any point in the process, but is not encouraged. Prompts consist of items that require an extensive amount of course content synthesis and application. Prior to the exam period, exam guidelines and rubrics are provided outlining the target requirements for content and degree of development necessary for maximum credit. Actual questions are not provided until the date of the exam. The intention of the task is to achieve a exam performance that is as close as possible to an applied behavioral performance as can be obtained with pen and paper.

Condition B: This format differs only in that groups submit only 1 set of responses as a collective, and therefore each receives the same grade.

2. Roundtable Interactive Peer Feedback Presentation Assessment:

Assessment Procedure: This presentation format varies from the traditional presentation in that students present their ideas to a series of smaller groups of peers in an interactive roundtable format as opposed to standing in front of the entire class and presenting with little or no interaction. Each roundtable session lasts about 15 minutes. Students are asked to provide a brief introduction and then peer groups are permitted to ask questions of the presenter. A rubric outlining what constitutes a quality presentation

is included in the course syllabus (Appendix C). Teacher assessment is obtained within one of the peer group sessions. In this session, the teacher is often required to ask questions that elicit evidence of both the content of the presentation as well as the students digestion of the critical issues related to their topic. Given that the presenters move from group to group, roughly the same amount of time is required as that for traditional presentations.

Study Methods

Participants consisted of students from 2 graduate education courses for each study condition (collaborative exam condition A: N=21, 25, condition B: N=122, 126; roundtable presentation N= 22, 23). Participants in all groups were surveyed after taking part in either of the respective assessment conditions. Surveys were constructed to obtain a measure of students' perceptions within each of the four dimensions of the construct for soundness. Following each exercise, volunteer were recruited for participation in focus group interviews. In these focus group interviews, 5-8 students were asked to discuss their experiences in more depth. For the collaborative exam condition B: focus group samples of 12 were selected for each section. Being that the participants for each condition consisted of the entire population of 2 required courses, the survey sample was considered fairly representative of all students admitted to these graduate certification programs. Moreover, the sample for the collaborative exams was obtained from universities in two separate geographical regions of the U.S.

RESULTS

Results from the survey and focus group data analysis (see data display below) showed findings that in some respects confirmed previous research, yet were surprising in other respects. In general, the collaborative method conditions received much better ratings than the traditional individualistic method conditions across all dimensions of soundness for both treatment groups. The only exception being that of the collaborative essay condition B which received higher marks on 3 of the 4 dimensions, falling below on reliability.

Initially, when considering implementing each of the conditions, researchers had little concern with their fundamental validity, but did question their ability to obtain reliable measures of performance. While participants had mixed feelings about the reliability of the collaborative exam, participants generally rated the reliability of both collaborative methods equal to or higher than traditional methods. This finding suggests that the primary concern for not using such practices, that students would feel that their grade was unfairly obtained, was not generally reported by these participants.

Possibly the most significant study findings for teacher educators were the participants' strongly positive feelings related to each assessment method's "effect on the learner." These findings supports previous research. For both methods, students felt strongly that it "promoted critical thinking," and "positive relationships among class members." For the roundtable method, participants overwhelming felt that it was "more enjoyable as an audience member," and "they learned more about the other members' presentation." For the collaborative exam, participants reported "learning more in the process," and being "more motivated to study." The fact that in a collaborative condition students tried harder is something of a surprise, given that many instructors would assume that students would take the opportunity to "ride on each others' coattails." From the

majority of accounts this was not the case with either group of participants. In fact, participants suggested they prepared more rigorously so that the would not "let their group mates down."

Date Display

Study data is displayed in this section 1) by survey mean for each of the four areas of soundness, then 2) with a representative sample of participant comments from the focus group interviews and survey comment sections, and finally 3) with the participant observations of the instructor. Survey means for reliability and validity are amalgamated from 3 items each. The efficiency rating, the effect on the learner rating, and overall soundness rating each reflect one item.

1. Reliability

Reliability – Roundtable: X= +0.4

```
+--------------+-------{}----|--------------+--------------+
Much Better      Better         Even           Worse          Much Worse
```

Participant Comments:

- *"It 's about the same [in response the question, do you think this format is as reliable?]"*
- *"Because [the instructor] could ask questions it made us have to be more prepared. I did not want to look stupid. If I just presented, I could talk about what I knew, but with the roundtable I had to be ready for people asking me hard implementation questions, so I had to be more prepared."*

Reliability – Collaborative Exam Condition A: X= +0.5

```
+--------------+-----{}-----|--------------+--------------+
Much Better      Better         Even           Worse          Much Worse
```

Participant Comments:

- Student generally agreed with the statement that this format could produce a reliable measure.
- Most students did not have a strong feeling one way or another. A few felt that they thought that theoretically there should be a lack of reliability but none expressed that they personally experienced a problem.

Reliability – Collaborative Exam Condition B: X=3

```
+--------------+-------------|-{}-----------+--------------+
Much Better      Better         Even           Worse          Much Worse
```

Participant Comments:

- *"I do not want to be mean, but there were a couple of people in my group that did not contribute at all."*
- *"I think it was reliable because it was a good way to see what we actually knew as opposed to a multiple-choice test like the midterm."*

- *"Some might have done all the work for the group."*

Instructor Participant Observation - Reliability

As the participants suggested, there was little difference in the reliability of the roundtables. In either case, the instructor would have used a clearly developed rubric (See Appendix C) and would be present for each presenter. The difference in the two cases would be the instructor's ability to ask questions and listen to group generated questions. This characteristic of the roundtable puts more control in the hands of the examiners and forces the presenter to defend and explain their ideas. In this sense, it could be suggested that there is generally greater reliability given the ability in this condition to determining what the presenter knows through something of a cross-examination.

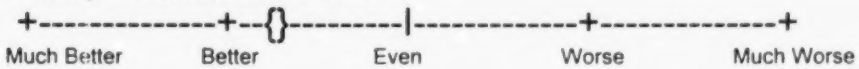
In the case of the collaborative exams, condition A demonstrated an unexpectedly good ability to determine the abilities of exam takers, and as expected condition B, showed little of such an ability to discriminate. Because groups turned in one set of responses, condition B fell prey to students who "rode the coattails" of their peers. However, in the cases where either all group members performed well, or performed poorly, the exam did provide a representative assessment of knowledge, preparation, or performance.

Nonetheless, in condition A, given the ability to assess individual papers independently, there was a fairly good ability to discriminate between the quality of each participants contribution. The responses of those who were more prepared were clearly distinguishable from those less prepared, in most cases. However, in groups where each member transcribed the group answer, members became indistinguishable. This can be reduced to some degree by instructions against using this strategy. Yet, overall, in condition A, students attempting to ride coattails or fake their way through were exposed pretty apparently.

In condition B, the area of reliability was a definite liability. It was impossible to discriminate one student's contribution from another. It was clear that there were some 20 percent of the students that were of little help to their group and may not have prepared to any great extent.

2. Validity

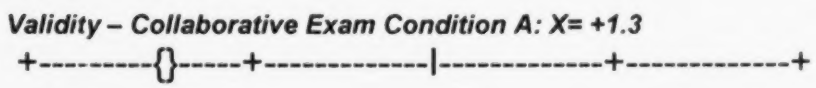
Validity – Roundtable: X +0.8



Participant Comments:

- *"I learned more about the projects. Asking questions enabled us to help the presenter with ideas or problems they had."*
- *"There was just more discussion and processing."*
- *"I still like the control of the (traditional) presentation."*

Validity – Collaborative Exam Condition A: X= +1.3



Much Better Better Even Worse Much Worse

Participant Comments:

- *"The state standards strongly suggest that teachers help create critical thinkers who can work well with others. I think that if teachers themselves do this, the students can model their behavior."*

- *"[This format] provides the real world experience of working as a team (teachers, T.A.s, Principals)."*

- *"Although the group and individual outcomes may both be valid, I think the individual on his/her own would arrive at a different solution if not influenced by group dynamics. The best way to assess an individual's knowledge is individually."*

Validity – Collaborative Exam Condition B: X= +0.9

+----------+{}----------|----------+----------+

Much Better Better Even Worse Much Worse

Participant Comments:

- *"We got to practice what we preach."*

- *"An understanding of the content was clearer."*

- *"At first I was hesitant about the exam, but after doing it I found that we could not have come up with a test this good alone. So I learned a lot and got a lot of encouragement for my ideas from the others. It was validating."*

- *"I am more comfortable doing things on my own – like I thought, just let me work by myself – this was uncomfortable. But I thought too that in real life you have to work with others like this and so I could see the value."*

Instructor Participant Observation - Validity

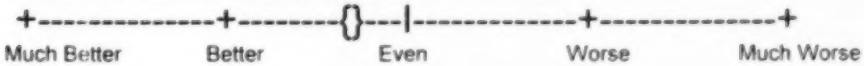
In all 3 cases, participants felt that a collaborative format was more valid than an individual format. This could be seen not only from the survey responses but from the verbal responses. Participants enthusiastically expressed their delight with the methods. As the comments suggest, participants found collaboration to be much more *authentic*. The roundtable format provided a venue to better process more complex aspects of the assignment than a stand up presentation. Participants felt that ideas in education are less often developed in a vacuum and more often are the result of collaborative discussion. The process could be observed to be more organic inasmuch as it was interactive and iterative. Products grew out of a generative process. This created a higher quality of product as well as a more satisfied producer.

In the exam conditions, students were generally surprised at what they found. They expected to have to compromise, which happened to some degree, but what they did not expect was how much better the quality of the ideas were that were ultimately generated. If they had guessed at their post-exam responses to the validity items they would not have been as high as they were. As students came up to turn in their exams they tended to be smiling. They felt very

accomplished, especially those who worked collaboratively with each item and did not divide the labor. And it should be noted that exams where students were more collaborative were better overall than those who reported having certain members focus more on certain sections and paste together a finished product.

3. Efficiency

Efficiency – Roundtable: X= +0.3



Participant Comments:

- *"Smaller groups. Questions from classmates promoted discussion."*
- *"It helped you write your paper (and with your idea) you could sit down with people and discuss it and find problems and get ideas so you could go back home and make changes."*
- *"I think the fact that we had started out the class working in cooperative groups helped make this work."*
- *"Maybe you could have a person designated as the facilitator for each session, that way you could keep people from wandering."*
- *"I think the fact that I missed a couple people still bugs me."*

Efficiency – Collaborative Exam (both conditions) X= +1.0

Participant Comments:

- *"[Great way to get] feedback on your ideas."*
- *"This is a great way to assess student's knowledge of material especially where there is so much material."*
- *"I think this format takes a certain amount of [discipline] I could see my 6th graders taking about everything but what they were supposed to be talking about at their roundtable when I was not at their table. And we did that too. . ."*
- *"If the class was not supportive like this one was, I don't know if I would have been comfortable doing this. I could not imagine presenting like this with the people in my high school [when I was a student]."*

Instructor Participant Observation - Efficiency

In each condition, the amount of work and coordination was about the same as that for the types of assessment with which they are being compared. The roundtable takes the same amount of time to do as regular presentations. The instructor gets the same total time with each participant

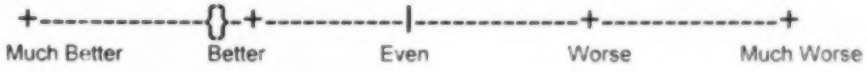
in the collaborative condition as they would in the traditional condition. But the fact that there are only 5-7 members in a group makes the opportunity to ask questions much more convenient. So with respect to getting at what the student knew, and for actually being of use in the thinking/writing process, the roundtable was more effective. The drawback is that no matter how one does the logistics, some students will not hear other student's presentations. In the end, students can hear the introduction to all the presentations, and can to take part in the roundtable portion for all but about 10-15 percent of their peers.

The collaborative exam condition A, where each student turned in a set of responses, is about the same logistically, after the exam, as if one had assigned the same essay items to individuals. Before the exam, there is a need to get students into groups and provide a set of study guidelines (see Appendix B), but this also has the benefit of structuring the exam preparation. So it is hard to tell if the amount of time is greater or lesser.

The primary reason that one would consider using the exam format in condition B, (having groups produce one set of responses per group), it would seem, has to do precisely with the issue of the efficiency or the shear quantity of work involved for the instructor. Clearly, reading a set of responses by a whole class of students is a lot of work. It takes about 10-30 minutes apiece to read exams completely. Making the choice between using collaborative exams and traditional essay exams with a manageable sized class did not pose any conflict between areas of soundness. However, assessing 120 student poses a dilemma. Assessing 120 sets of essay responses is unreasonable whether they were completed within a collaborative format or an independent format. So the choice is to do a collaborative exam where groups turn in one set of answers (producing about 25 exams to grade), or to give an objective test. In this case, the choice was based on the notion that soundness would be best served if a collaborative exam were used, knowing that reliability was the price for gaining the other benefits desired.

4. Effect on Learner

Effect on Learner - Roundtable Aggregate X=1.2



Survey Item Means:

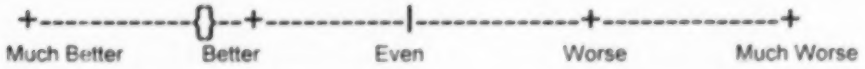
Enjoyed as an audience member	+1.6
Promoted positive relationships	+1.5
Caused more critically thinking	+1.2
Learned more about other presentations	+1.2
Helped in writing process	+0.5
Motivational	+0.3

Participant Comments:

- *"Smaller groups. Questions from classmates promoted discussion."*
- *"Held small audience better. Had to respond to Q and A you might not have thought of."*
- *"Socially I think you would get to know people better."*

- *"[if you have an interactive mechanism] It helps you think about your topic better."*
- *"I liked the familiarity of this format over the other, because it promoted a different mindset."*
- *"I could not imagine doing this with a class that was not supportive like this one was. If it was a hostile class, then I can't imagine. . ."*

Effect on Learner - Collaborative Exam (both conditions aggregate) X=1.2



Survey Item Means:

Promoted positive relationships	+1.6
Caused more critically thinking	+1.6
Learned more in process	+1.0
Motivational	+0.5

Participant Comments:

- *"Helped me think the questions out more, explain my thinking and therefore clarify my answers more."*
- *"Ownership of the material (peer pressure) more likely to be prepared in order to not let the group down."*
- *"The process reinforced my confidence in my knowledge of the content."*
- *"Fosters teamwork. Allows for peer teaching."*
- *"Exchange of ideas. Reminder of things/concepts learned, but temporarily forgotten."*
- *"The material was discussed, debated, and then written, allowing students to develop a deeper understanding."*
- *"Helped me understand how to do a very worthwhile alternative assessment method."*
- *"[This format promoted many] levels of skills, cognition, organization- group is bigger than sum of its parts."*
- *"6 months down the road if you tested us again, I think we would know the material better after going through this process. I really think we will remember it better."*
- *"It lets you know how the children feel when you ask them to work collaboratively."*
- *"The [exam] seemed secondary to the feelings I got working with the group.".*

Instructor Participant Observation – Effect on Learner

The most notable observation regarding how the collaborative conditions benefited the students was that they did not foresee beforehand how the process would effect them. Before the exam took place, most students were either mildly optimistic or somewhat indifferent to the thought of being assessed using a collaborative structure, but a good number were uncomfortable with the idea. This discomfort seemed to be most related to the methods being "different" and odd, and also that they required one to work outside of his/her comfort zone, especially in the case of the collaborative exam. It was not uncommon to hear questions such as "why are we doing it this way?" or "I don't see the purpose of doing this." But, in most cases this attitude changed after they took part in the activity. It was not uncommon to hear the comment after the exam, "I did not think this was going to work, but it really did help me ___." Not all students were sold on the idea after taking part in the assessment, but as the survey data suggests, they walked away with a very positive impression of what they had done. I would guess that if this survey was given to the participants before they had done it, and if they were asked to predict their feelings about the methods, they would not have expressed nearly as positive attitudes toward the idea of working collaboratively.

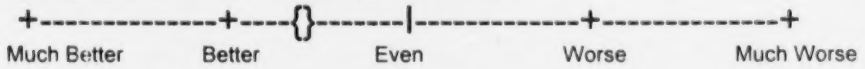
The best analogy I can find to characterize most students' feelings after completing the collaborative exam (each condition), is that of being part of a "winning team." Succeeding as part of a team, it could be said, may be more satisfying than succeeding as an individual. Participants typically expressed a very vivid sense of accomplishment after completing the collaborative exam. This observation reflects what could be seen as a stand-alone benefit of using such a system, but it may also explain the homogeneously positive rating most participants typically gave to the collaborative condition in general. That is to suggest, the feeling of "winning" may potentially have influenced the objectivity of participants on their survey ratings.

In terms of the motivational influence, the roundtable appeared to be more motivational due to the sense of accountability and responsibility. The collaborative exam also seemed to be more motivational to most in each condition. But there were a very few in condition A that "slacked" a bit (maybe 5%) because they knew the others in their group would be prepared. However in condition B, there were maybe 20-30 percent that did not prepare as rigorously. An observation that was made by this instructor and many students was that one the one hand, a collaborative outcome is motivating to students with a high sense of group responsibility and on the other hand, it can be an opportunity to ride on the coattails of the better prepared for students with a low sense of group responsibility.

5. Overall Soundness Rating

Participant Survey Ratings for Overall Soundness:

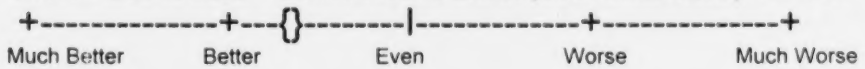
Overall Soundness - Roundtable X= +0.6



Summary of Overall Survey Results – Roundtable Presentations

Reliability = Not a significant concern (as might have been expected).
Validity = More authentic.
Helped in writing process.
More engaging and educational for audience.
Efficiency = About the same with improvement suggestions.
Benefits = Students worked just as hard or harder.
Promoted more collegial environment.
Promoted higher levels of critical thinking.

Overall Soundness – Collaborative Exam (both conditions) X= +0.7



Summary of Overall Survey Results – Collaborative Exams

Reliability = A hypothetical concern of some, but not tangibly experienced by participants in condition A. Inability to detect "slackers" was a significant problem in condition B.
Validity = More authentic given nature of teacher work.
Efficiency = No real difference.
Benefits = Students worked just as hard or harder.
Promoted better interpersonal relationships.
Promoted higher levels of critical thinking.

An interesting result from the collaborative exam data was that participants, in all 4 sample groups, as well as the participant observer's experience of the two conditions was almost identical for all 4 dimensions of soundness. Whether participants were ultimately responsible for their own responses or if they contributed to a single collective effort they reported a similar set of experiences after the exam.

STUDY LIMITATIONS

1. The effectiveness of either of these assessment conditions may need to be examined within the context of their use. A great deal of collaborative work was incorporated into each of these classes before the assessments took place. Additionally, grading in each course was characterizes within a cooperative and criterion-referenced orientation. The results of this study may not be easily generalized to less cooperative and/or norm-referenced structured courses.

2. The instructor's assessment skills and/or relationship with the class may be factors in the perceived effectiveness of either method. This may be especially true of the area of reliability. If the participants' were not able to *trust* the instructor's ability to objectively apply the prescribed criteria, and/or they were suspicious of the instructor's intentions for using such methods, the reported ratings for reliability (and possibly for all 4 areas of soundness) may not have been as high.

3. As discussed earlier, the emotions related to a sense of group accomplishment or "winning" were still fresh in the minds of participants as they completed the surveys and took part in the focus groups. This positive emotion could have been associated with the collaborative methods. And while this may have had a desirable effect on learning, it may have colored their ratings of some of the technical aspects of the process they were evaluating.

4. The focus group interviewer/moderator was also the instructor of course. Students may have edited themselves to some degree as a result. There was no cost to honesty, but some participants may have edited their feelings. Likewise, some degree of "expectancy" could have been reflected as well.

DISCUSSION

The findings of this study suggest that, in the hands of an instructor who is committed to cooperative learning, has creating clear and well-established targets, and is trusted by her/his students, it appears collaborative assessments have the potential to achieve a high degree of soundness. In fact, in this limited study, participants did not seem to see much if any of the downside that critics might have anticipated. The collaborative exams did not seem to be any more trouble or any less "fair." Yet, beyond fears related to logistics and "fairness," there seems to be an upside to collaborative assessment that may not be able to be achieved by other forms of assessment.

The question that I ask myself (in my role as a responsible instructor) after gathering data from six classes at two universities and talking to students formally and informally about their thoughts and feelings, is simply, "knowing what I know, should I keep assessing with collaborative methods?" My answer would be that even if I had to pay a price in the area of reliability or efficiency, which in most cases I do not feel I did, I would make every effort to incorporate collaborative assessment. I can think of four main reasons why I have come to this conclusion.

First, Students seem to learn more in collaborative conditions. As participants suggested, working with others promotes a type of thinking that seems to be more critical and longer lasting. As opposed to being limited to the thoughts in one's own mind, which in many cases are flawed or by definition restricted, the student can incorporate a broader and inherently more diverse set of ideas. Therefore, what is ultimately constructed in developing a response to an exam, or the ideas being examined in a roundtable discussion, are of higher quality. And as the ideas take form on paper and in the memories of the students, they are more thoughtful and well-conceived.

Second, I liked what I observed the collaborative assessment conditions promoting. Personally, I do not want to promote learning as a form of transmission and retention. Too often our students in teacher education are what Carol Dweck (2000) calls "helpless pattern" thinkers, who are more interested in getting answers right than growing as learners. I see this all too clearly, and feel that I have to do everything I can to help them work out of what Dweck calls a "mastery orientation." I feel, if they practice thinking of success more as taking advantage of the opportunities within the learning condition, and not so much just getting right answers, they will be less inclined to promote that thinking in their students. Collaborative assessment seems to provide a great capacity to promote a constructivist epistemological foundation in a course. Moreover, I like that collaborative assessment, along with collaborative learning activities, promotes an atmosphere in a class that supports risk taking and an environment where and a sense of community can develop. This atmosphere just does not happen unless students are required to invest in one another in a meaningful and substantive manner.

Third, where else do students learn to sink or swim in a collective effort? If we withhold this experience of mutual interdependence we are denying our students one of a limited number of opportunities to develop these critical skills. I recall the focus group participant who lamented that in her first year of teaching she struggled, but did not know how to work with others or to come out of herself to get the help that she needed. She realized it was her mindset in which she saw herself as all alone that kept her isolated. We in teacher education talk a lot about the value of working collaboratively, but we stop short of actually creating learning environments where we force our students to move outside of their comfort zones and give up independent control over their learning. These are skills that members of well-functioning teams learn. As was depicted in the findings, maybe the most enduring aspect of the experience of taking part in a collaborative assessment was the sense that one's team "won."

Fourth as participants suggested, working in an interdependent condition is in their minds, closer to what the job of teaching *should* look like. While most pre-teachers do not see collaboration in the schools they come in contact with to the degree that they feel it should be present, they felt that "good teaching" is inherently collaborative. There is a great deal of research to support them in this contention. Therefore, if any practice can achieve something close to an authentic experience of teaching, we have some obligation to find ways to incorporate it on a practical level. Where else in a student's college experience do they learn to work as a team? And as a growing body of research is showing, students in teacher preparation programs reproduce to a great extent the pedagogical methods that were used in their programs. Promoting such sound and generative practice is even more salient in the field of education, due to the propensity of students to model the practices to which they have been exposed, and to determine the legitimacy of a practice by its use or non-use by the "experts."

These practices are not for everyone. There is a sincere commitment to the value of collaboration required. Moreover, early efforts to incorporate collaborative assessment will likely feel uncomfortable and odd. Students who have experienced year after year, and course after course of individual assessments will resist the notion of working with others with a meaningful outcome on the line. The data displayed here represent assessments that took place at the end of courses where substantive collaboration had been used regularly and purposefully. It is not evident whether such methods would be

experienced as soundly by either the students or the instructor, if the groundwork for the relational and technical context had not already been set. But it appears from these data that thinking about assessment collaboratively does not inherently lead one down the road to pedagogy that is structurally deficient. In fact, if assessment is viewed within a broader domain of "soundness" which includes consideration for its "effect on the learner," assessment done without the benefit of collaboration can appear lacking in some ways. As has been found in previous examinations of collaborative exams (Stearns, 1996; Yuretich, Khan, & Leckie, 2001) there is a processing that occurs within a group that can not occur in the mind and experience of an isolated individual. Thus the level of critical thinking, retention, and sense of accomplishment may only be possible within a collaborative context. Likewise, without a collaborative element to presentations the depth of processing of the presenter, and the engagement and level of learning by the participants may be less achievable without a collaborative component.

The results of this study, as well as the limited number of studies before it examining collaborative assessment, suggest that there are few downsides and potentially significant upsides to the use of such practices. These finding would indicate that more attention and further research is warrented into this area.

CONCLUSION

The results of this study suggest that each of these forms of collaborative assessment *can* be accomplished in ways that are *sound*. It therefore affords teacher educators the legitimacy necessary to incorporate these useful techniques for promoting the crucial, albeit difficult to teach, skills and dispositions related to collaboration in their students.. Given the increasing language related to promoting collaboration skills within the standards documentation from major professional teacher education organizations, there appears to be a growing awareness of the critical place collaboration plays in good teaching. It could be said that good teaching has always needed to be collaborative, and collegiality continues to be a requisite condition for a highly functioning school (Glickman, 1993; Hargreaves, 1994). The mandate is evident that we as a field must find ways to foster these skills and dispositions in our students. Because if we in teacher education want our candidates to approach their own work and their students' learning with the necessary emphasis given to collaboration, we much provide the experiential learning context by providing for meaningful use of collaborative practices.

REFERENCES

Antony, J. (1994) *Defining the Teaching-Learning Function in Terms of Cooperative Pedagogy: An Empirical Taxonomy of Faculty Practices.* Paper presented at the Annual Meeting of the Association for the Study of Higher Education. Tucson, AZ, November 10-13)

Bohde, C. (1996) New Teacher Collaboration. Creating an Integrated Biology/English Unit. *The Science Teacher* v. 63 pp. 28-31.

Brown, S. & Glasner, A. (1999) *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches.* Society for Research into Higher Education Ltd. London.

Buckelew, M. (1991) *Group Discussion Strategies for a Diverse Student Population.* Paper presented at the Annual Meeting of the Conference on College composition and Communication. Boston March 21-23.

Dweck, C.S. (1999) *Self-Theories: Their role in motivation, personality, and development.* Taylor and Francis, Lillington, NC.

Fullan, M. (1993) *Change Forces: Probing the Depths of Educational Reform.* Falmer Press, Toronto.

Glickman, C. D. (1993) *Renewing America's schools: A guide for school-based action.* San Francisco: Josses-Bass.

Hargreaves, A. (1994). *Changing teachers, changing times.* New York: Teachers College Press.

Hermann, C. (1995) Creative Group Presentations in Organic Chemistry. *Journal of Chemical Education*, v.72, p.157.

Hooper S. (1992) *Generative Learning in Small Groups.* In Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology. (ERIC Reproduction Service Number ED347995).

McDonald, H. (1989) Small Group Oral Presentations in Historical Geology. *Journal of Geological Education* v. 37 pp. 49-52.

National Council for Accreditation of Teacher Education (1997). *Technology and the New Professional Teacher – Preparing the 21st Century.* National Council for Accreditation of Teacher Education Task Force on Technology.

Schumm, J. (1995) An Assessment Alternative: Group Oral Exams. *Journal of Reading* v. 38 p.490.

Shindler, J., Yang, H., Nephew, J., & Keen. (2000) *Examining the Soundness of Process-Based Assessment within an Applied Technology* Course. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28.

Stearns, S. (1996) Collaborative Exams as Learning Tools. *College Teaching* v.44 pp. 111-12.

Yuretich, R., Khan, S. & Leckie, R. (2001) Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course. *Journal of Geoscience Education*, v.49, no. 2, pp. 111-19

Appendix A: Working Definition of Assessment Soundness.
The following definition of soundness was provide students during the focus group interviews:

Validity:
- Assessment measures what it intends to measure
- Assessment measures the most relevant learning from course/assignment content
- Assessment method is well matched to the assessment target

Reliability:
- Assessment device could be used reliably by two different individuals
- Assessment device could be used reliably for repeated trials/performances
- An appropriate sample of performances is collected to represent a true representation of performance/ability
- Performance criteria is described in measurable, specific, concrete, objective outcome terms

Efficiency:
- Assessment data can be collected in an efficient, timely, doable manner
- Assessment does not unnecessarily interfere with teaching or learning tasks

Influence on Student Affect:
- Assessment procedure has an overall positive affect on the *student-teacher relationship*
- Assessment has an overall positive affect on the student's *motivation* level
- Assessment promotes a sense of *competence* by providing +/- performance feedback
- Assessment creates a sense of *internal locus of control* by providing a clear and attainable target and path to attaining it.
- Assessment creates a greater sense of *belonging and cooperation* among the members of the class.

Appendix B: Incorporating Collaborative Exams into a Course:

Step 1: Prepare students for the material to be covered, have students work in groups for previous cooperative activities, and then let student select groups of 3-5 (with the option of working alone).

Step 2: Provide guidelines for what should be in a quality response. This enables students to prepare more purposefully. Example of guidelines for one of 3 items assigned.

Collaborative Exam Study Outline

For this exam you will be able to work in groups of 3-5 (your choice of members). You will be allowed to bring with you 2 pages of notes in addition to this sheet, but other than these notes this is a closed book exam. You will be given the entire period for the exam. Each member of the group will submit and be assessed on only their particular exam (there is no expectation papers will agree). The exam is worth 40 points, and will consist of the following 3 essay question concepts.

Essay Item #1

Given a learning outcome that you are asked to (hypothetically) teach to a class, design a strategy to accomplish the learning. Lead your reader through what you would do with the students and your planning thought process.

An excellent response would include the following:

- Use of lesson planning language appropriate to your methodological strategy (This can come from any source you choose).
- Demonstration of a good understanding of matching instructional strategies with content goals.
    - does the lesson require some form of direct instruction (if it needs to be modeled and practiced it probably does)
    - would the lesson be more effective if the students discovered the principles on their own (inductively)
    - is there a concept involved (if so how is that concept going to be developed)?
    - do you want students to make judgments or reflect on ideas?
- Inclusion of some learning activities that you would consider most effective.
    - does the outcome lend itself to cooperative learning?
    - does the lesson need an advanced organizer (book, activity, concept map)?
    - how will you know the students are getting it, or not?(generally address assessment)
    - talk generally about how the students will accomplish the learning

Step 3: Provide directions and test items the day of the exam. For example the following: (verbal directions should accompany written directions)

Final Exam

Answer the following questions on separate sheets of paper. Each item is worth 10 points. Responses should be sufficiently developed (per the exam review guidelines), and will likely require at least two pages of elaboration. Your hypothetical class can be any grade level you choose.

NOTE: You may talk with others, but all responses need to be prepared independently, your answers should be your own.

1. Pat is eating a pear at lunch. Chris looking on remembers when he had pear trees at his house in Idaho. He recalled that the pears all fell off the trees by October. Since its April, he makes the comment to Pat, "Where did that pear come from?" (You have noticed the Nicaragua stickers on the pears you have been seeing in the stores lately) They are both perplexed. They return from lunch and ask you in front of the whole class. You may not be sure of the answer, but you chose to take the opportunity to teach the concept of as part of your science instruction. Specifically describe how you would go about teaching this concept. As much as possible, try to metacognitively walk the reader through your thinking and instructional decision making.

Appendix C: Incorporating a Roundtable Presentation:

Step 1: create a rubric for a quality presentation. For example the following:

Roundtable Presentation Guidelines

	Topic Explanation	Implementation	Visuals
Excellent	**6 pts.** Topic is clearly explained and well defined. Problem/need is identified. Significance is addressed. Goal is stated.	**5 pts.** A well-conceived implementation plan is evident from discussion. Solutions to problematic areas of implementation have been considered. Evidence of assessment strategy.	**4 pts.** Visuals are used purposefully to aid in the understanding of the topic. Key concepts are represented.
Good Effort	**5 pts.** Topic is explained and defined. Problem/need is identified. Goal/need is evident. Significance is evident.	**4 pts.** An implementation plan is evident from discussion. Problematic areas of implementation are considered. Evidence of assessment strategy.	**3 pts.** Visuals are used to aid in the understanding of the topic.
Adequate	**3-4 pts.** Topic is explained. Problem/need is identified.	**2-3 pts.** An implementation plan is evident from discussion.	**2 pts.** Visuals are used.
Problematic	**0-2 pts.** Topic is discussed	**0-1 pts.** Implementation has been considered.	**0 pts.** No visuals

Step 2: Provide directions before the students present. It may be useful to include one portion of the presentation that is done in front of the whole class as an advanced organizer and introduction as outlined in the following directions:

Presentation Guidelines and Assessment

Project presentations will be done in a roundtable format. This format will be discussed in more detail in class, but presenters will have @5 minutes to present individually in front of the whole class. In that time, there should be an attempt to give the audience a general idea of the project including:

- Purpose of the study/project
- Problem statement
- Need determination/communication procedure
- Context of study or project

After that presenters will have @10-20 minutes with a series of 2-5 groups to discuss their ideas and implementation. In that time, the presenter will have the opportunity to discuss their ideas in more depth. Implementation, leadership, project development can all be dealt with here. Plan to take about half the time to talk and about half to respond to the groups questions.

Step 3: Instructor stays with one roundtable group while presenters switch groups for each session. The instructor should take the opportunity to ask questions that can delve into areas of understanding and digestion.

Step 4: Instructor should provide a written assessment immediately following the presentation. Using the rubric with a space for comments works well.

Step 5: It is most desirable if students have the opportunity to make revisions to their written work, incorporating ideas and feedback from the roundtable, before it is collected.

BEST COPY AVAILABLE